

PERANCANGAN SISTEM PENGENALAN SUARA DENGAN METODE LINEAR PREDICTIVE CODING

SPEECH RECOGNITION SYSTEM DESIGN USING LINEAR PREDICTIVE CODING METHOD

R Ricki Juniansyah¹, Ir. Rita Magdalena, M.T.², Ledy Novamizanti, S.Si., M.T.³
^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom
¹rickijuniansyah@students.telkomuniversity.ac.id, ²ritamagdalenat@telkomuniversity.ac.id,
³ledyaldn@telkomuniversity.ac.id

Abstrak

Speech recognition merupakan teknologi untuk mengenali suara manusia yang kemudian dapat dimanfaatkan untuk berbagai keperluan. Pada tugas akhir ini *speech recognition* digunakan untuk merancang sistem aplikasi pengenalan suara menggunakan metode *Linear Predictive Coding*.

Linear Predictive Coding merupakan suatu metode analisa sinyal yang menghasilkan sejumlah koefisien LPC. Pada tugas akhir ini dilakukan ekstraksi ciri suara dengan metode LPC dan dilanjutkan dengan metode *Hidden Markov Model* untuk pelatihan dan pengenalan suara. LPC mengambil ciri-ciri yang ada didalam sampel suara manusia kemudian menjadi masukan untuk pembelajaran pola. Pada penelitian sebelumnya metode LPC digunakan dalam sistem *speech to text*.

Hasil pengujian sistem pada tugas akhir ini menunjukkan bahwa akurasi tertinggi sistem bernilai 53,34% dengan waktu komputasi 51,27 detik. Hasil tersebut didapat pada kondisi ukuran frame bernilai N=900 dan M=700, 7 jumlah data training tiap *user*, ukuran *codebook* 4 dan jumlah *state* 7.

Kata Kunci : Linear Predictive Coding, Hidden Markov Model

Abstract

Speech recognition is a technology to recognize human voices which can be used for various purposes. In this final project, speech recognition is used for design system using Linear Predictive Coding Method.

Linear Predictive Coding is an analysis method which generates a signal LPC coefficients. In this final project, Linear Predictive coding is used for feature extraction, then for pattern recognition will be using Hidden Markov Model method. LPC generates the characteristics in samples of human voice and then it will be input to the Hidden Markov Model learning process. In previous research LPC method is used for speech to text system design.

Testing results of the system in this final project shows that the highest accuracy at 53,34% with 51,27 seconds of computing time. The results obtained at conditions of N=900 and M=700 for frame blocking, 7 samples data per user for training, 4 codebook size, and 7 state.

Keywords: Linear Predictive Coding, Hidden Markov Model

1. Pendahuluan

Teknologi pengenalan suara saat ini telah mengalami perkembangan yang cukup pesat. Banyak aplikasi-aplikasi pengenalan suara yang telah dihasilkan seiring dengan berkembangnya teknologi. *Speech Recognition* merupakan teknologi yang memanfaatkan sinyal suara manusia sebagai masukan untuk dikenali oleh sistem, yang kemudian dapat digunakan untuk berbagai keperluan [10]. Sinyal suara akan diproses dan diambil ekstraksi cirinya sehingga menghasilkan sejumlah informasi yang dapat dianalisis.

Untuk membuat sistem pengenalan suara dibutuhkan beberapa metode agar memperoleh hasil yang baik. Pada tugas akhir ini akan dibuat sebuah sistem pengenalan suara dengan metode *Linear Predictive Coding* untuk ekstraksi ciri, dan metode *Hidden Markov Model* untuk pembelajaran pola sehingga suara dapat dikenali.

Linear Predictive Coding (LPC) merupakan teknik analisis sinyal suara yang telah teruji dan sangat baik dalam proses ekstraksi ciri suara [12]. LPC memproses sinyal suara yang kemudian menghasilkan *feature* (ciri) suara yang akan diproses selanjutnya. *Hidden Markov Model* (HMM) adalah sebuah metode statistika yang sangat baik untuk diterapkan pada sistem pengenalan suara [7]. *Feature* (ciri) suara akan menjadi masukan metode HMM dan melatih pola dari *feature* tersebut sehingga suara dapat dikenali. Pada tugas akhir ini dibuat sebuah sistem untuk mengenali suara manusia sebagai *password*.

2. Dasar Teori

2.1 Sinyal Suara[4][5]

Sinyal dapat diklasifikasikan menjadi beberapa jenis yaitu : sinyal waktu kontinyu, sinyal waktu diskrit, sinyal nilai kontinyu, sinyal nilai diskrit, sinyal random, dan sinyal nonrandom. Sinyal waktu kontinyu dengan nama lain sinyal analog adalah sinyal yang belum melalui proses apapun. Sedangkan sinyal nilai diskrit atau sinyal digital adalah sinyal analog yang telah melalui proses *sampling*, *quantization*, dan *encoding*.

2.2 Linear Predictive Coding[1][8]

Linear Predictive Coding merupakan teknik analisis sinyal percakapan yang menyediakan ekstraksi fitur yang berkualitas baik dan efisien untuk digunakan dalam perhitungan. LPC pertama kali digunakan pada tahun 1978 untuk membuat alat sintesis sinyal percakapan.

Ide dasar dari LPC model adalah sinyal suara pada saat waktu n dilambangkan dengan $x(n)$, yang dapat diperkirakan dengan menggunakan kombinasi linier dari sampel suara sebelumnya.

$$x(n) \approx a_1 x(n-1) + a_2 x(n-2) \dots + a_p x(n-p) \quad (2.1)$$

Dimana koefisien a_1, a_2, \dots, a_p merupakan koefisien dari LPC, dan p orde dari LPC.

2.3 Vector Quantization[8]

Vector Quantization diperlukan untuk menghasilkan deretan observasi (indeks *codebook*) yang selanjutnya akan diproses menggunakan metode *Hidden Markov Model* (HMM) untuk proses training.

Codebook adalah sekumpulan titik (*vector*) yang mewakili representasi sinyal suara dalam bentuk digital dalam ruang suara. Untuk kasus pengenalan suara, maka *codebook* ini harus dibuat untuk suara yang akan dikenali dengan cara menggunakan metode *clustering* terhadap fitur suara. Dengan *clustering* yang dilakukan akan membuat suara yang memiliki kemiripan akan berada dalam satu kelompok yang sama.

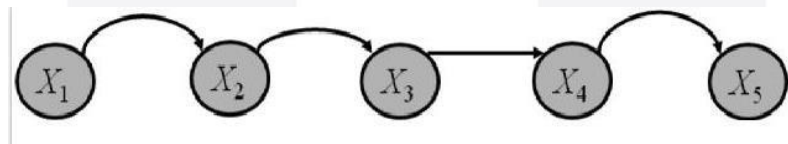
Ada berbagai jenis algoritma *clustering* yang ada, pada tugas akhir ini algoritma *clustering* yang digunakan adalah *K-Means clustering*

2.4 Hidden Markov Model[6][7]

Hidden Markov Model (HMM) adalah sebuah metode statistika yang sangat baik untuk diterapkan pada sistem pengenalan suara. HMM memiliki simbol keluaran atau fungsi peluang yang menggambarkan simbol keluaran berhubungan dengan *state* atau transisi antar *state*. Observasi tiap *state* digambarkan secara terpisah dengan suatu fungsi probabilitas atau fungsi densitas yang didefinisikan sebagai peluang untuk menghasilkan simbol tertentu saat terjadi transisi antar *state*.

Algoritma HMM didasari oleh model matematik yang dikenal dengan rantai Markov. Beberapa hal yang dapat dijelaskan tentang rantai *Markov* yaitu:

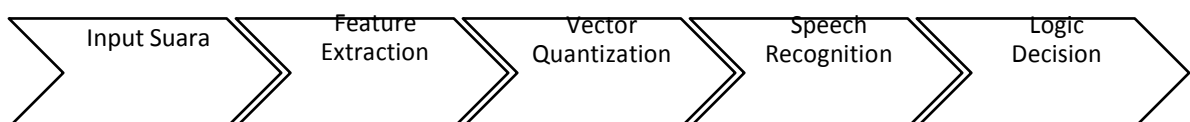
1. Transisi Keadaan dari suatu keadaan tergantung pada keadaan sebelumnya.
2. Transisi keadaan bebas terhadap waktu.
3. Rantai *Markov* secara umum ditunjukan pada gambar



Gambar 2.6 Rantai *Markov*

3. Perancangan dan Implementasi

Pada tugas akhir ini akan dibangun sebuah sistem pengenalan suara dengan menggunakan *Linear Predictive Coding* (LPC) sebagai metode ekstraksi ciri suara dan *Hidden Markov Model* (HMM) sebagai metode *classifier* atau pembelajaran pola suara pada proses *training* dan *testing*. Secara garis besar, desain model sistem ini terdiri dari dua proses yaitu proses *training* dan proses *testing*.



4. Pengujian dan Analisis

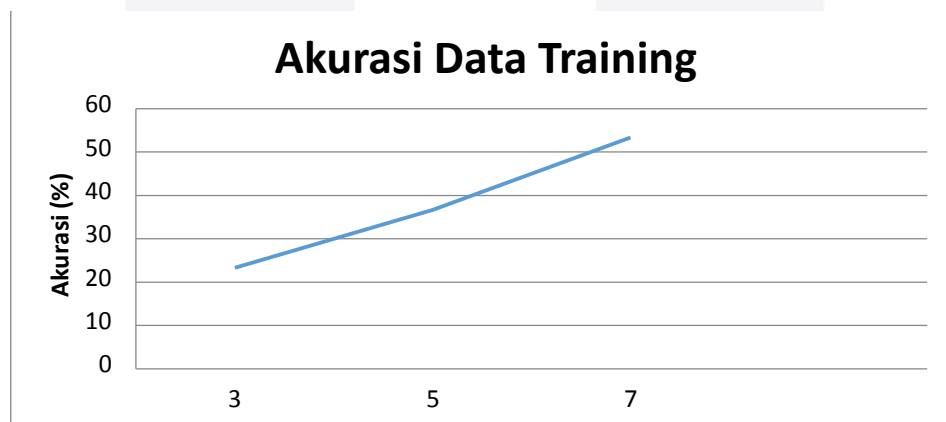
Sistem yang telah dirancang perlu diketahui akurasi dengan beberapa pengujian. Bab ini akan membahas skenario yang digunakan untuk menguji tingkat akurasi dari sistem yang telah dibuat serta menyajikan hasil-hasil pengujiannya.

4.1 Skenario Pertama

a. Tujuan Pengukuran

Pada skenario pertama akan dilihat bagaimana pengaruh jumlah data training dari tiap-tiap user terhadap tingkat akurasi dan waktu komputasi sistem. Pengujian dilakukan pada jumlah data training dari tiap user sebanyak 3, 5, dan 7

b. Hasil Pengukuran



Gambar 4.1 Grafik Tingkat Akurasi Pada Skenario 1

c. Analisis Hasil Pengukuran

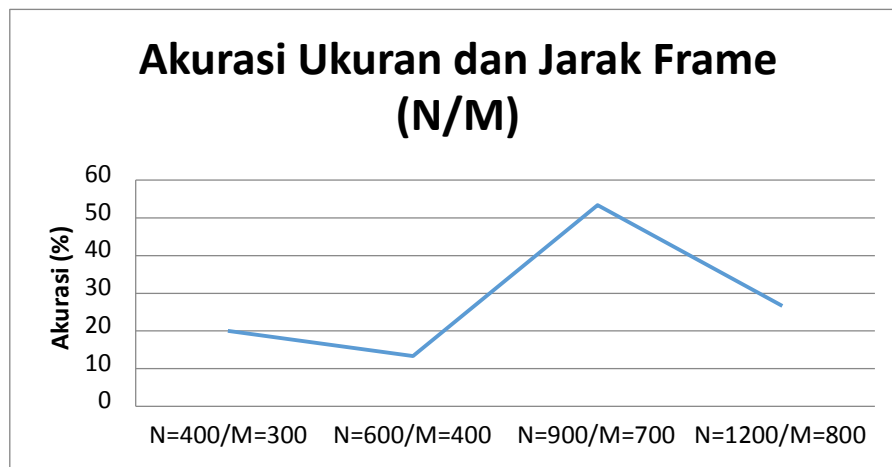
Dari grafik di atas terlihat bahwa jumlah data *training* yang banyak menghasilkan tingkat akurasi yang jauh lebih baik dan stabil dibandingkan dengan sampel data yang sedikit. Namun jumlah sampel data yang banyak juga mempengaruhi waktu komputasi sistem. Semakin banyak jumlah sampel data *training* maka semakin lama waktu komputasi dari sistem.

4.2 Skenario Kedua

a. Tujuan Pengukuran

Skenario kedua dilakukan untuk melihat pengaruh ukuran (N) dan jarak *overlap frame* (M) pada proses *frame blocking* terhadap tingkat akurasi dan waktu komputasi sistem.

b. Hasil Pengukuran



Gambar 4.2 Grafik Tingkat Akurasi Pada Skenario 2

c. Analisis Hasil Pengukuran

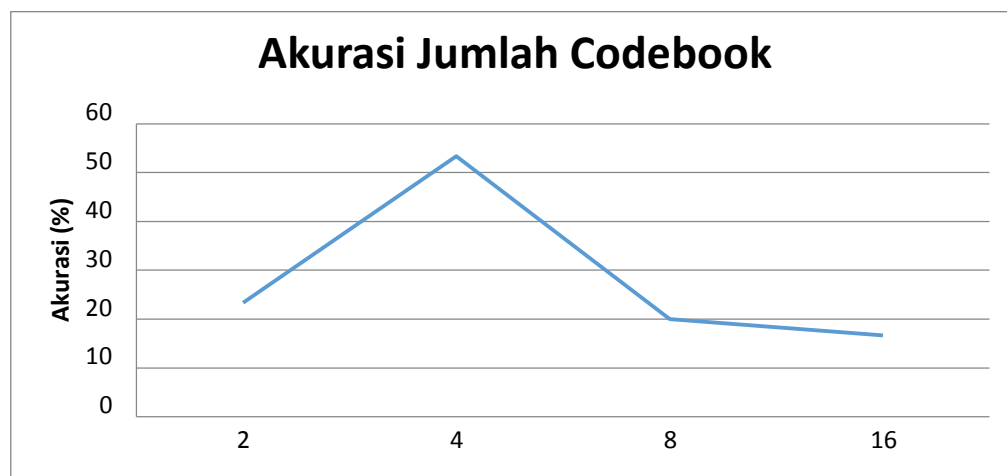
Dari grafik di atas terlihat bahwa ukuran *frame* mempengaruhi tingkat akurasi sistem. Semakin kecil ukuran dan jarak *frame* maka akan semakin banyak jumlah *frame* yang akan dihasilkan pada proses *frame blocking*, maka semakin sensitif juga sistem saat mengenali suara. Namun sistem yang terlalu sensitif membuat tingkat akurasi menurun karena menyebabkan terlalu banyaknya jumlah *frame* dan ciri dari suara membuat sistem kesulitan untuk mengenali suara. Pada pengujian ini diperoleh ukuran dan jarak *frame* yang memiliki akurasi tertinggi sebesar 53.34% yaitu dengan ukuran *frame* (N) = 900 sampel dan jarak *overlap* antar *frame* (M) = 700 sampel.

4.3 Skenario Ketiga

a. Tujuan Pengukuran

Skenario ketiga dilakukan untuk melihat pengaruh jumlah *codebook* terhadap tingkat akurasi dan waktu komputasi sistem. Pengujian ini dilakukan dengan jumlah *codebook* 2, 4, 8, dan 16.

b. Hasil Pengukuran



Gambar 4.3 Grafik Tingkat Akurasi Pada Skenario 3

c. Analisis Hasil Pengukuran

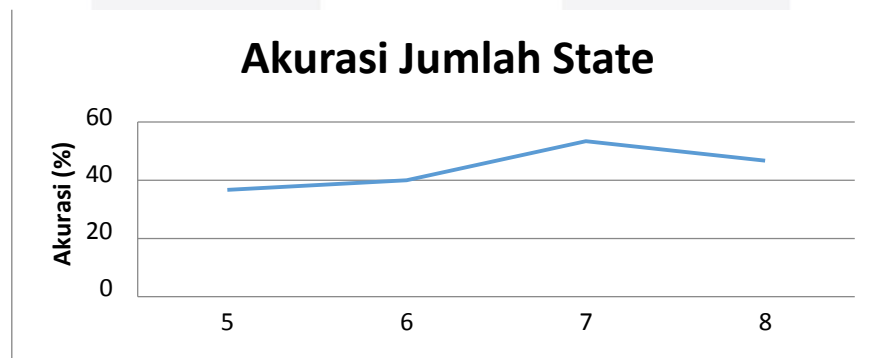
Dari grafik di atas terlihat bahwa akurasi terbaik terdapat pada jumlah *codebook* sebanyak 4 dengan tingkat akurasi sebesar 53,34%. Dari analisa yang dilakukan dapat diketahui bahwa jika jumlah *codebook* terlalu banyak terdapat kemungkinan bahwa terdapat sekumpulan titik vektor yang seharusnya berada pada satu *cluster*, namun karena terlalu banyaknya batasan titik tersebut menjadi terpisah. Hal ini menyebabkan sistem menjadi sangat sensitif dan sulit untuk mengenali suara. Penentuan jumlah *codebook* yang tepat sangat penting untuk memperoleh tingkat akurasi yang baik tergantung dari kondisi dan tipe suara yang akan dikenali sistem.

4.4 Skenario Keempat

a. Tujuan Pengukuran

Skenario keempat dilakukan untuk melihat pengaruh jumlah *state* pada pembelajaran pola HMM terhadap tingkat akurasi dan waktu komputasi sistem. Pengujian ini dilakukan menggunakan jumlah *state* 5, 6, 7, dan 8.

b. Hasil Pengukuran



Gambar 4.4 Grafik Tingkat Akurasi Pada Skenario 4

c. Analisis Hasil Pengukuran

Berdasarkan grafik diatas dapat dilihat bahwa setiap jumlah *state* memiliki tingkat akurasi dan waktu komputasi yang hamper sama. Oleh karena itu dapat disimpulkan bahwa jumlah *state* tidak terlalu berpengaruh baik pada akurasi maupun terhadap waktu komputasi dari sistem.

5.1 Kesimpulan

Dari hasil pengujian dan analisis simulasi tugas akhir ini, dihasilkan beberapa kesimpulan sebagai berikut :

1. Sistem pengenalan suara sebagai *password* yang telah dibangun yang kemudian dilakukan pengujian dengan beberapa skenario memperoleh tingkat akurasi terbaik sebesar 53,34% dengan waktu komputasi 51,27 detik.
2. Jumlah sampel data *training* tiap-tiap user yang paling baik digunakan untuk sistem ini adalah 7 data sampel. Karena jumlah sampel tersebut menghasilkan tingkat akurasi sebesar 53,34% dengan waktu komputasi selama 51,27 detik.
3. Ukuran dan jarak *overlap* antar *frame* (nilai N dan M) yang paling baik digunakan untuk sistem ini adalah N=900 sampel dan M=700 sampel. Karena ukuran dan jarak *overlap frame* tersebut menghasilkan tingkat akurasi rata-rata sebesar 53,34%.
4. Jumlah *codebook* yang paling baik digunakan untuk sistem ini adalah 4 *codebook*. Karena jumlah *codebook* tersebut menghasilkan tingkat akurasi sebesar 53,34% dengan waktu komputasi selama 51,27 detik.
5. Peningkatan jumlah *state* tidak selalu membuat tingkat akurasi sistem semakin besar, dengan kata lain jumlah *state* tidak terlalu berpengaruh terhadap tingkat akurasi maupun waktu komputasi dari sistem.
6. Cara pengambilan sampel suara dan pemilihan jumlah data sampel suara memiliki pengaruh terhadap tingkat akurasi dan waktu komputasi sistem.

DAFTAR PUSTAKA

- [1] Thiang and Suryo. Wijoyo, "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling movement of Mobile Robot", Petra Christian University (2011)
- [2] Irfandy. Mahmud, "Aplikasi Pengenalan Ucapan Dengan Jaringan Syaraf Tiruan Propagasi Balik Untuk Pengendalian Robot Bergerak", Universitas Diponegoro (2010)
- [3] Roberts, K., Stodden, R "The Use of Voice Recognition Software as a Compensatory Strategy for Postsecondary Education Students Receiving Services Under the Category of Learning Disabled", Journal of Vocational Rehabilitation (2005)
- [4] Lawrence Rabiner, and Bing Hwang Juang, "Fundamentals of Speech Recognition", Prentice Hall, New Jersey. (1993)
- [5] Picone, Joseph. "Fundamental of Speech Recognition", Mississippi State University (1996)
- [6] L.E Baum and T. Petrie, "Statistical Inference for Probabilistic Function of Finite State Markov Chains", Ann Math. Stat, vol 37, (1996)
- [7] Prasetyo. B Eko, "Teori Dasar Hidden Markov Model", Institut Teknologi Bandung (2010)
- [8] Devesh. Nema, K. Raghavan, and A. Sandeep, "Implementation of Speech Recognition in Resource Constrained Environments", Indian Institute of Technology Roorkee. (2006)
- [9] Dave. Namrata, "Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition", International Journal for Advance Research in Engineering and Technology (2013)
- [10] Fauzi, R.M, Adiwijaya, Maharani W. "The Recognition of Hijaiyah Letter Pronunciation using Mel Frequency Cepstral Coefficients and Hidden Markov Model", International Journal Conference on Data, Internet, Education & Technologies (2016)
- [11] Aidil F. Muhammad, "Speech Recognition dengan Ekstraksi Fitur Linear Predictive Coding dan JST Cerebellar Model Articulation Controller", Universitas Telkom (2011)
- [12] H. Fandy, Suyanto T. Iwan, "Pengenalan Sinyal Suara Pada Speech to Text Menggunakan Linear Predictive Coding dan Hidden Markov Model" Universitas Telkom (2011)